Diffusion, dimensionality and noise in transcriptional regulation

Gašper Tkačik and William Bialek[*]

Joseph Henry Laboratories of Physics, Lewis–Sigler Institute for Integrative Genomics,
and [*]Princeton Center for Theoretical Physics, Princeton University, Princeton, New Jersey 08544
Center for Studies in Physics and Biology, Rockefeller University,
1230 York Avenue, New York, New York 10065
(Dated: February 5, 2008)

The precision of biochemical signaling is limited by randomness in the diffusive arrival of molecules at their targets. For proteins binding to the specific sites on the DNA and regulating transcription, the ability of the proteins to diffuse in one dimension by sliding along the length of the DNA, in addition to their diffusion in bulk solution, would seem to generate a larger target for DNA binding, consequently reducing the noise in the occupancy of the regulatory site. Here we show that this effect is largely cancelled by the enhanced temporal correlations in one dimensional diffusion. With realistic parameters, sliding along DNA has surprisingly little effect on the physical limits to the precision of transcriptional regulation.

I. INTRODUCTION

Cells constantly regulate the expression levels of their genes. A central motif in this regulatory process is the binding of transcription factor proteins to specific sites along the DNA. The precision of transcriptional regulation is limited, ultimately, by randomness in the arrival of transcription factor (TF) molecules at these sites [1]. But proteins can find their binding sites on DNA by two very different mechanisms—either by diffusing in three dimensions through the surrounding solution, or by binding weakly and diffusing in one dimension along the contour of the DNA molecule. The idea of diffusion or sliding along DNA goes back (at least) to the realization that the *lac* repressor seems to bind more rapidly to its target site than would be allowed by three dimensional diffusion alone [2, 3]. More recently, the discussion of 3D vs. 1D diffusion has been revitalized by theoretical analysis of the optimal search strategies [4], by new biochemical measurements [5], and by direct physical observations of the sliding motion [6, 7, 8]. Here we consider the impact of dimensionality and diffusion on the physical limits to the precision of transcriptional regulation.

Any physical system which responds to the concentration of a signaling molecule will exhibit noise due to the random diffusion of these molecules in and out of its "sensitive volume" [1]; the larger the volume, the smaller the (fractional) noise. This leads to the intuition that 1D diffusion will have a huge impact on the noise because it effectively increases the size of the target to which the transcription factor is binding [9]. Our main result is that this intuition is wrong. The problem is that diffusion in one dimension has a very different statistical structure than in three dimensions, and so one cannot simply say that 1D sliding generates a larger 3D target. We show that realistic combinations of 3D and 1D diffusion in fact have a surprisingly small effect on the limiting noise level—the increased size of the target is largely cancelled by stronger temporal correlations, which means that integrating over a fixed time

gives fewer independent samples.

II. CONCENTRATION FLUCTUATIONS

Before starting on the problem of proteins binding to DNA, it is useful to recall some facts about concentration fluctuations for molecules free in solution [10]. If we write the concentration as $c(\mathbf{x}, t)$, then the power spectrum of fluctuations $S_c(\mathbf{k}, \omega)$ is defined by

$$
\langle \delta c(\mathbf{x}, t) \delta c(\mathbf{x}', t') \rangle
$$
$$
= \int \frac{d^d \mathbf{k}}{(2\pi)^d} \int \frac{d\omega}{2\pi} S_c(\mathbf{k}, \omega) e^{+i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')-i\omega(t-t')}, \quad (1)
$$

where d is the dimensionality of space. For molecules present at the mean concentration $\bar{c}$ and diffusing freely with diffusion constant D, the power spectrum is given by

$$
S_c(\mathbf{k}, \omega) = \frac{2\bar{c}Dk^2}{(Dk^2)^2 + \omega^2}, \quad (2)
$$

where $k = |\mathbf{k}|$. This result is independent of the dimensionality d, although of course the units of concentration are different in different dimensions. If we integrate over all frequencies, which corresponds to making instantaneous measurements, we find a spatial power spectrum

$$
\int \frac{d\omega}{2\pi} S_c(\mathbf{k}, \omega) = \bar{c}. \quad (3)
$$

This is spatial white noise, and embodies the fact that a snapshot of molecules in solution will reveal a random, Poisson distribution: the variance of the molecule number in any finite volume is equal to the mean, so the power spectrum of concentration fluctuations is equal to the mean concentration, and again this is independent of dimensionality.

Consider a measurement of concentration that is averaged over some small region of space,

$$
C(t) = \int d^d \mathbf{x}\, W(\mathbf{x}) c(\mathbf{x}, t), \quad (4)
$$

where W is a weighting or windowing function that defines the size of the region we are interested in. To interpret C as an average of the local variable c, we must have $\int d^d\mathbf{x} W(\mathbf{x}) = 1$. The temporal fluctuations in C are themselves determined by a power spectrum $S_C(\omega)$,

$$\langle \delta C(t)\delta C(t')\rangle = \int \frac{d\omega}{2\pi} S_C(\omega) e^{-i\omega(t-t')} \tag{5}$$

$$S_C(\omega) = \int \frac{d^d\mathbf{k}}{(2\pi)^d} S_c(\mathbf{k},\omega)|\tilde{W}(\mathbf{k})|^2, \tag{6}$$

where

$$\tilde{W}(\mathbf{k}) = \int d^d\mathbf{x}\, W(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{7}$$

The normalization of W implies that $\tilde{W}(\mathbf{k}=0) = 1$. If the size of the region is ℓ^d, then $\tilde{W}(\mathbf{k})$ will fall to zero for k much larger than $\Lambda \sim 1/\ell$. Thus we can approximate

$$S_C(\omega) = \int \frac{d^d\mathbf{k}}{(2\pi)^d} \frac{2\bar{c}Dk^2}{(Dk^2)^2+\omega^2}|\tilde{W}(\mathbf{k})|^2 \tag{8}$$

$$\sim \int_0^\Lambda dk\, k^{d-1} \frac{2\bar{c}Dk^2}{(Dk^2)^2+\omega^2}. \tag{9}$$

We are interested in the behavior of the power spectrum at low frequencies, corresponding to measurements with long averaging times. If we try simply to set $\omega = 0$, then we find

$$S_C(\omega=0) \sim \int_0^\Lambda dk\, k^{d-1} \frac{2\bar{c}}{Dk^2}. \tag{10}$$

For $d=3$,

$$S_C^{3d}(\omega=0) \sim \int_0^\Lambda dk\, k^2 \frac{2\bar{c}}{Dk^2} \sim \frac{2\bar{c}\Lambda}{D} \sim \frac{\bar{c}}{D\ell}. \tag{11}$$

This corresponds to white noise in the time domain, so we expect that averaging over time will reduce the noise variance in proportion to the averaging time. More precisely, if we average for a time τ_{int}, we will be sensitive to frequencies $|\omega| < 1/\tau_{\text{int}}$, so we will see a variance

$$\langle(\delta C)^2\rangle_{3d} \sim \int_{|\omega|<1/\tau_{\text{int}}} \frac{d\omega}{2\pi} S_C^{3d}(\omega) \tag{12}$$

$$\sim \int_{|\omega|<1/\tau_{\text{int}}} \frac{d\omega}{2\pi} S_C^{3d}(\omega=0) \tag{13}$$

$$\sim \frac{\bar{c}}{D\ell\tau_{\text{int}}}. \tag{14}$$

Although we have not done a full calculation including the effects of binding and unbinding to target sites, this is essentially the "noise floor" for a system which senses the concentration of signaling molecules by having them bind to a site of size ℓ [1, 11, 12].

In contrast, for $d=1$ we have

$$S_C^{1d}(\omega=0) \sim \int_0^\Lambda dk\, \frac{2\bar{c}}{Dk^2}, \tag{15}$$

which is infrared divergent. To get the right low frequency behavior in one dimension we have to be a bit more careful. We have

$$S_C^{1d}(\omega) \sim \int_0^\Lambda dk \frac{2\bar{c}Dk^2}{(Dk^2)^2+\omega^2}. \tag{16}$$

This integral is actually finite as $\Lambda \to \infty$, so we can write

$$S_C^{1d}(\omega) \sim \int_0^\infty dk \frac{2\bar{c}Dk^2}{(Dk^2)^2+\omega^2} \tag{17}$$

$$= \frac{2\bar{c}}{\sqrt{D\omega}} \int_0^\infty d\left(k\sqrt{\frac{D}{\omega}}\right) \frac{Dk^2/\omega}{(Dk^2/\omega)^2+1} \tag{18}$$

$$= \frac{2\bar{c}}{\sqrt{D\omega}} \int_0^\infty dq \frac{q^2}{q^4+1} \tag{19}$$

$$\sim \frac{\bar{c}}{\sqrt{D\omega}}. \tag{20}$$

Now the variance of measurements averaged over a time τ_{int} becomes

$$\langle(\delta C)^2\rangle_{1d} \sim \int_{|\omega|<1/\tau_{\text{int}}} \frac{d\omega}{2\pi} S_C^{1d}(\omega) \tag{21}$$

$$\sim \int_{|\omega|<1/\tau_{\text{int}}} \frac{d\omega}{2\pi} \frac{\bar{c}}{\sqrt{D\omega}} \tag{22}$$

$$\sim \frac{\bar{c}}{\sqrt{D\tau_{\text{int}}}}. \tag{23}$$

We see that in one dimension, the *variance* of concentration measurements declines only as the square root of the measurement time. This is contrast to the three dimensional case, where the standard deviation of the noise declines as the square root of time but the variance declines in direct proportion to the measurement time, as seen in Eq (14).

III. BINDING WITH 3D DIFFUSION

If transcription factors bind specific sites on the DNA to control the rate of gene expression, then the noise in the regulated gene product will contain a contribution from the noise in the occupancy of the specific site. Here we want to compute this noise contribution, starting with the case where the transcription factors find their cognate site by 3D diffusion alone and progressing towards more complicated translocation strategies in the following sections.

In this section we briefly review the calculations of Ref [11]. Consider a cell volume with an average concentration $\bar{c}$ of TFs present, and a binding site located at $\mathbf{x}_0 = 0$. The transcription factor molecules can diffuse with a bulk diffusion constant D_3, bind to the specific site at a rate k_+c, and dissociate back into the free solution at a rate k_-, according to the following set of

equations:

$$\frac{\partial c}{\partial t} = D_3 \nabla^2 c - \dot{n}\,\delta(\mathbf{x} - \mathbf{x}_0), \qquad (24)$$

$$\dot{n} = k_+ c(\mathbf{x}_0, t)(1 - n) - k_- n. \qquad (25)$$

To determine the fluctuations in the binding site occupancy n, we linearize the equations about the mean occupancy $n(t) = \delta n(t) + \bar{n}$ with $\bar{n} = k_+ \bar{c}/(k_+ \bar{c} + k_-)$, and around the mean concentration, $c(\mathbf{x}, t) = \delta c(\mathbf{x}, t) + \bar{c}$, and write the perturbations as Fourier modes:

$$\delta n(t) = \int \frac{d\omega}{(2\pi)} e^{-i\omega t}\, \delta\tilde{n}(\omega), \qquad (26)$$

$$\delta c(\mathbf{x}, t) = \int \frac{d\omega}{(2\pi)} \int \frac{d^3 k}{(2\pi)^3} e^{-i(\omega t + \mathbf{k}\cdot\mathbf{x})}\, \delta\tilde{c}_{\mathbf{k}}(\omega). \quad (27)$$

We are looking for the power spectrum of fluctuations in occupancy, which we can define by

$$\langle \delta\tilde{n}(\omega)\delta\tilde{n}^*(\omega') \rangle = 2\pi\delta(\omega - \omega') S_n(\omega). \qquad (28)$$

If there exists a low frequency limit, $S_n(\omega \to 0)$, and we average the noise for a time τ_{int} that is longer than all other relevant timescales in the problem, we can [by analogy with the arguments leading to Eq (14)] use it to calculate the observable variance by computing $\sigma_n^2 = S_n(\omega \to 0)/\tau_{\text{int}}$.

Having introduced the necessary notation, we can compute the fluctuations in concentration *at* the binding site, $\mathbf{x}_0 = 0$, from Eq (24):

$$\delta\tilde{c}(\mathbf{x}_0, \omega) = i\omega\delta\tilde{n} \int \frac{d^3 k}{(2\pi)^3} \frac{1}{-i\omega + D_3 k^2}. \qquad (29)$$

This integral is ultraviolet divergent, and must be cut off at $\Lambda = \frac{\pi}{a}$, where a is the binding site size; keeping leading terms in $i\omega$ one obtains:

$$\delta\tilde{c}(\mathbf{x}_0, \omega \to 0) = \frac{i\omega\delta\tilde{n}}{2\pi D_3 a}. \qquad (30)$$

The linearization of Eq (25) yields

$$-i\omega\delta\tilde{n} = -\frac{\delta\tilde{n}}{\tau_c} + k_+(1 - \bar{n})\delta\tilde{c}(\mathbf{x}_0, \omega) + k_-\bar{n}\beta\delta\tilde{F}, \quad (31)$$

where $\tau_c^{-1} = k_+ \bar{c} + k_-$ is the time scale of occupancy fluctuations, $\beta = 1/k_B T$, and $\delta\tilde{F}$ is the fluctuation in the free energy difference between the bound and empty states of the site; it is the 'force' that is thermodynamically conjugate to the 'displacement' δn. After substituting the local concentration fluctuations, Eq (30), into Eq (31), one can finally use the fluctuation–dissipation theorem to compute the power spectrum of the noise in occupancy:

$$S_n(\omega) = \frac{2k_B T}{\omega} \mathcal{I}m\, \frac{\delta\tilde{n}}{\delta\tilde{F}}. \qquad (32)$$

The result of Ref [11] was that, in the case of 3D diffusion to the binding site, the low-frequency limit of the occupancy noise power spectrum is:

$$S_n(\omega \to 0) = \frac{2\bar{n}(1 - \bar{n})^2}{k_-} + \frac{\bar{n}^2(1 - \bar{n})^2}{\pi D_3 a\bar{c}}. \qquad (33)$$

The first term corresponds to the binomial switching fluctuations as the occupancy of the specific site changes between full and empty; this term depends on the microscopic details of the TF-DNA interaction (here embodied by the off-rate constant k_-). The second term is caused by the fluctuating diffusive flux to the binding site, and provides a lower bound on the noise, independent of details. Expressed in terms of the equivalent concentration fluctuations, $S_c(\omega) = S_n(\omega)\left|\frac{d\bar{n}}{dc}\right|^{-2}$, the noise spectrum for the second term is

$$S_c(\omega \to 0) = \bar{c}/\pi D_3 a, \qquad (34)$$

with the associated fractional variance

$$\left(\frac{\sigma_c}{\bar{c}}\right)^2 = \frac{1}{\pi} \times \frac{1}{D_3 \bar{c}\tau_{\text{int}}} \times \frac{1}{a}, \qquad (35)$$

where, to facilitate comparison with results that follow, we explicitly pull out the leading numerical factor and $1/(\text{effective binding site size})$. This is the noise lower bound, consistent with the result of Eq (14) for $\ell = a$, that depends solely on the rate of 3D diffusion D_3, the binding site size a, and the amount of time averaging that the system performs, τ_{int}.

IV. BINDING WITH 1D DIFFUSION

Consider the case where the diffusion in bulk does not occur, and the TF is only free to slide along the length of the DNA. Let $\xi(x, t)$ denote the 1D concentration along the DNA contour, which we imagine to be stretched in the $\hat{\mathbf{x}}$ direction, and let D_1 be the corresponding diffusion constant. Equations (24,25) remain unchanged but for the new notation, i.e. $c(\mathbf{x}, t) \to \xi(x, t), D_3 \to D_1, \nabla^2 \to \partial_x^2$; we will use q to denote the Fourier variable conjugate to x. We again need to calculate the concentration fluctuations at the binding site (cf. Eq (29)):

$$\delta\tilde{\xi}(x_0, \omega) = i\omega\delta\tilde{n} \int_{-\infty}^{\infty} \frac{dq}{(2\pi)} \frac{1}{-i\omega + D_1 q^2}. \qquad (36)$$

This integral diverges as $q \to 0$ if we set $\omega = 0$, but in contrast to the discussion above there is no problem as $q \to \infty$. The full result is that:

$$\delta\tilde{\xi}(x_0, \omega) = \frac{i\omega\delta\tilde{n}}{2D_1\sqrt{-i\omega/D_1}}. \qquad (37)$$

Note that unlike Eq (30), this result does not contain the binding site dimensions. Using Eq (37) we compute

the diffusive contribution to the noise power spectrum (cf. Eq (34), but no limit $\omega \to 0$):

$$S_\xi(\omega) = \frac{\bar{\xi}}{\sqrt{2\omega D_1}}. \tag{38}$$

The noise variance is obtained by integrating the noise power spectrum, as in the arguments leading to Eq (23),

$$\left(\frac{\sigma_\xi}{\bar{\xi}}\right)^2 = \frac{2}{\pi} \frac{1}{\sqrt{2D_1 \tau_{\rm int} \bar{\xi}}}. \tag{39}$$

We see that the noise variance declines as $1/\sqrt{\tau_{\rm int}}$, consistent with the simpler calculation of concentration fluctuations in Eq (23). This again is in contrast to our intuition that variances should decline as $1/\tau_{\rm int}$, which is the result for binding coupled to 3D diffusion. The difference between 1D and 3D can be understood by realizing that in 1D diffusion, a particle leaving the binding site at the origin has a large probability (in fact, probability 1) of returning back to the origin. A receptor trying to estimate the local concentration of TFs will therefore be unable to collect samples that are truly independent, and the variance in the measurements will consequently decrease at a rate that is slower than expected from averaging over independent measurements.

Suppose that the DNA with the sliding TFs is embedded into the cytoplasm, where the bulk concentration of TFs is $\bar{c}$; the TFs can jump onto the DNA at a rate (per unit length) $\kappa_+ \bar{c}$, where we expect $\kappa_+ \sim D_3$. The TF will stay on the strand for a *residence time*, $\tau_{\rm r} = \kappa_-^{-1}$, where κ_- is the rate for dissociating from the DNA. There will be an equilibrium between the 1D and 3D concentrations, $\bar{c} = \bar{\xi}/\tau_{\rm r} \kappa_+ \approx \bar{\xi}/\tau_{\rm r} D_3$. With these identifications, the noise variance can be rewritten in a form similar to the result in Eq (35):

$$\left(\frac{\sigma_c}{\bar{c}}\right)^2 = \frac{\sqrt{2}}{\pi} \times \frac{1}{D_3 \bar{c} \tau_{\rm int}} \times \frac{1}{\sqrt{D_1 \tau_{\rm r}^2 / \tau_{\rm int}}}. \tag{40}$$

Here, the effective binding site size is $a_{\rm eff} = \sqrt{D_1 \tau_{\rm r}} \sqrt{\tau_{\rm r}/\tau_{\rm int}}$. Naively, one could think that the effective site size would be equal to the average length that the TF explores during 1D diffusion on the DNA, $b = \sqrt{D_1 \tau_{\rm r}}$, and therefore $a_{\rm eff} \sim b$, but that would be wrong: the effective site size depends on $\tau_{\rm int}$ to compensate for the highly correlated fluctuations in 1D diffusion. Because $\tau_{\rm int} \gg \tau_r$ [13], the effective site size will be much smaller than b. The apparent reduction in the noise expected in the naive picture – because the size of the binding site, a, is replaced by presumably much larger length explored by 1D diffusion, b – must in reality be traded off against the longer required integration time. As a result, it is not immediately clear whether 1D diffusion decreases the noise lower bound compared to 3D case.

V. COMBINED 1D AND 3D DIFFUSION

In the discussion so far we have been missing a parameter that would interpolate between two qualitatively different noise regimes: the 3D result of Eq (35) and the 1D result of Eq (40). If the rate for dissociation from the sliding state into the bulk is increased, the 1D case must approach the 3D result; the residence time will grow ever shorter and will, at some scale, break the strong correlations reflective of the 1D sliding mode with its high probability for returning to the origin.

A. Perfect mixing in bulk solution

We first solve a simplified case where the TF diffuses on the DNA, but if it dissociates into the bulk, it mixes very quickly and there is no correlation between the point of dissociation and subsequent re-association to the DNA strand. We expect that the divergence observed in the pure 1D case, Eq (36), will now get regularized. The
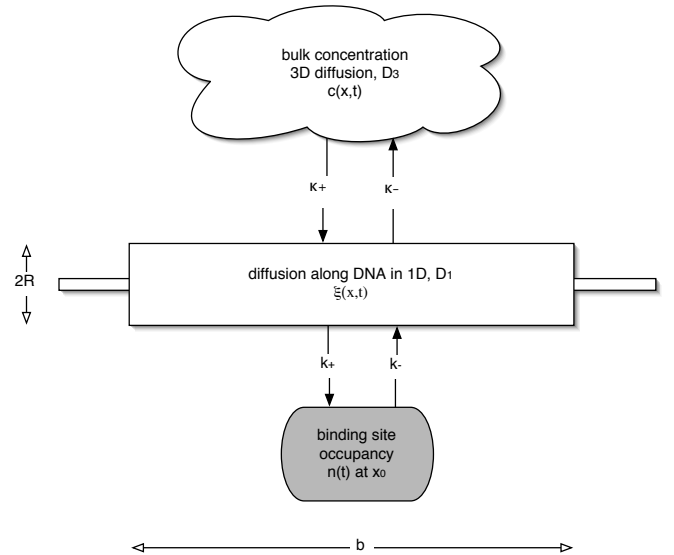

FIG. 1: The transcription factors can either be free in solution at concentration c, or they can enter a region on the DNA where they diffuse by sliding. The 1D concentration is denoted by $\xi(x, t)$. The specific binding site on the DNA is at location $x_0 = 0$; $k_+ \xi(x_0)$ and k_- are the on– and off–rates for transitions between the 1D sliding state and the bound state on the specific site; $\kappa_+ c$ is the rate per unit length for transition onto the DNA from free solution, and $\kappa_- = \tau_{\rm r}^{-1}$ is the rate for dissociating from 1D solution into the 3D solution. The effective radius of the DNA molecule is R and the "sliding length," or the average distance along the contour covered in a single 1D random walk before dissociation, is denoted by b.

dynamical equations for this case are:

$$\frac{dn}{dt} = k_+\xi(x_0,t)(1-n) - k_-n \tag{41}$$

$$\frac{\partial\xi}{\partial t} = D_1\frac{\partial^2\xi}{\partial x^2} - \dot{n}\delta(x-x_0) - \kappa_-\xi + \kappa_+C \tag{42}$$

$$\frac{dC}{dt} = \kappa_-\int dx\,\bar{\xi}(x,t) - \kappa_+LC \tag{43}$$

Here, $C(t)$ is the bulk concentration that is perfectly mixed and therefore does not have any spatial dependence; TFs in the bulk are enclosed into a box with a side of length L and the DNA inside is an infinitely thin line of length L; and κ_- and κ_+ are rates for transitioning between 1D concentration ξ and the bulk C. When linearized and written out in terms of their Fourier components, the equations couple to the zero mode, $\delta\xi(q=0,\omega)$, differently than to non-zero modes:

$$(-i\omega + 1/\tau_c)\delta\tilde{n} = k_+(1-\bar{n})\delta\tilde{\xi}(x_0) + k_-\bar{n}\beta\delta F \tag{44}$$

$$(-i\omega + D_1q^2)\delta\tilde{\xi}_q = i\omega\delta\tilde{n} - \kappa_-\delta\tilde{\xi}_q + \kappa_+L\delta\tilde{C} \tag{45}$$

$$(-i\omega + \kappa_+L)\delta\tilde{C} = \kappa_-\delta\tilde{\xi}_{q=0} \tag{46}$$

We use Eq (45) at $q=0$ together with Eq (46) to express $\delta\tilde{C}$ and substitute it back into Eq (45) for non-zero q; it is then straightforward to write $\delta\tilde{\xi}(x_0)$ as:

$$\delta\xi(x_0) = \int\frac{dq}{(2\pi)}\frac{i\omega\Sigma(\omega)\,\delta\tilde{n}}{-i\omega + D_1q^2 + \kappa_-} \tag{47}$$

$$= \frac{i\omega\Sigma(\omega)}{2\sqrt{D_1}}\frac{\delta\tilde{n}}{\sqrt{\kappa_- - i\omega}}, \tag{48}$$

where

$$\Sigma(\omega) = 1 + \frac{\kappa_-\kappa_+L}{(-i\omega + \kappa_-)(-i\omega + \kappa_+L) - \kappa_-\kappa_+L}. \tag{49}$$

Note that as $\omega\to 0$, $\Sigma(\omega)$ will diverge; before computing the power spectrum of fluctuations S_n using Eq (32) we

therefore need to expand Eq (47) to the leading power of $i\omega$. On the other hand, the momentum integral in Eq (47) now converges due to the κ_- term in the denominator.

After some algebra, the effective power spectrum in the concentration fluctuations is:

$$S_\xi(\omega\to 0) = \frac{\bar{\xi}}{2\sqrt{D_1\kappa_-}}\left(1 + \frac{1-\gamma}{(1+\gamma)^2}\right), \tag{50}$$

where $\gamma = \kappa_+L/\kappa_-$ is the ratio of on- and off-rates for going into 1D solution. The γ-dependent expression in parenthesis has a minimum of $7/8$ at $\gamma_0 = 3$, and therefore a noise floor exists that cannot be eliminated by some convenient combination of the rate parameters.

Rewritten in terms of the "equivalent concentration" $\bar{c} = \bar{C}/L^3 = \kappa_-\bar{\xi}/\kappa_+L^3 \approx \kappa_-\bar{\xi}/D_3$, the concentration noise floor is (cf. Eq (40)):

$$\left(\frac{\sigma_c}{\bar{c}}\right)^2 = \frac{7}{16}\times\frac{1}{D_3\bar{c}\tau_{\mathrm{int}}}\times\frac{1}{\sqrt{D_1\tau_{\mathrm{r}}}}. \tag{51}$$

This is the result we might naively have expected, namely that sliding along the DNA creates a larger target, of size $a_{\mathrm{eff}}\approx b$. But we get this result *only* because we have assumed that the TF can repeatedly dissociate into a well-mixed bulk solution, instantly losing the memory of the point along the DNA from which it dissociated.

B. Diffusion in bulk solution

We will now solve the coupled 3D-1D diffusion problem without assuming that in the bulk solution the transcription factors mix perfectly. The relevant quantities are schematized in Fig 1.

We describe the system by the following set of equations:

$$\frac{dn}{dt} = k_+\xi(x_0,t)(1-n) - k_-n, \tag{52}$$

$$\frac{\partial\xi(x,t)}{\partial t} = D_1\frac{\partial^2\xi(x,t)}{\partial^2 x} - \frac{dn}{dt}\delta(x-x_0) + \kappa_+c(x,\mathbf{R}=0,t) - \kappa_-\xi(x,t), \tag{53}$$

$$\frac{\partial c(x,\mathbf{R},t)}{\partial t} = D_3\nabla^2 c(x,\mathbf{R},t) - [\kappa_+c(x,\mathbf{R}=0,t) - \kappa_-\xi(x,t)]\,\delta(y)\delta(z). \tag{54}$$

We have again assumed that DNA is stretched along the x-axis and that it is an infinitely thin molecule. ξ is a function of only one variable, x, while c is a function of coordinate x and radial coordinates $\mathbf{R}$ that are perpendicular to the $\hat{\mathbf{x}}$ direction.

We linearize and Fourier transform the equations as follows:

$$-i\omega\delta\tilde{n} = -\frac{\delta\tilde{n}}{\tau_c} + k_+(1-\bar{n})\delta\tilde{\xi}(x_0) + k_-\bar{n}\beta\delta\tilde{F} \tag{55}$$

$$-i\omega\delta\tilde{\xi}_q = -D_1q^2\delta\tilde{\xi}_q + i\omega\delta\tilde{n} - \kappa_-\delta\tilde{\xi}_q + \kappa_+\delta\tilde{c}_q \tag{56}$$

$$-i\omega\delta\tilde{c}_{q,\mathbf{k}_\perp} = -D_3\left(q^2 + k_\perp^2\right)\delta\tilde{c}_{q,\mathbf{k}_\perp} + \kappa_-\delta\tilde{\xi}_q - \kappa_+\delta\hat{c}_q, \tag{57}$$

where again q is the spatial Fourier variable conjugate to x and $\mathbf{k}_\perp$ is conjugate to $\mathbf{R}$. Note that $\delta\tilde{\xi}_q$ is function only of q, while $\delta\tilde{c}_{q,\mathbf{k}_\perp}$ is a 3D Fourier transform of the bulk concentration fluctuations that depends on both q and $\mathbf{k}_\perp$. Finally, $\delta\hat{c}(q,\mathbf{R}=0,\omega)$ are Fourier modes of concentration fluctuations along the $\hat{x}$ direction, evaluated at the location of the DNA strand, $\mathbf{R}=0$.

We first compute $\delta\hat{c}_q$ from Eq (57):

$$\delta\hat{c}_q = \int \frac{d^2k_\perp}{(2\pi)^2}\delta\hat{c}_{q,\mathbf{k}_\perp} \tag{58}$$

$$\delta\hat{c}_q = \int \frac{d^2k_\perp}{(2\pi)^2}\frac{\kappa_-\delta\hat{\xi}_q - \kappa_+\delta\hat{c}_q}{-i\omega + D_3q^2 + D_3k_\perp^2} \tag{59}$$

$$\delta\hat{c}_q = \frac{\kappa_-\delta\hat{\xi}_q - \kappa_+\delta\hat{c}_q}{4\pi D_3}\log\left\{1+\frac{\Lambda^2}{k_0^2}\right\} \tag{60}$$

Here, $k_0^2 = q^2 - i\omega/D_3$, and $\Lambda = \frac{\pi}{R}$ is the ultraviolet cutoff at the radial size of the DNA molecule. We can substitute $\delta\hat{c}_q$ from Eq (60) into Eq (56) to obtain the expression for $\delta\hat{\xi}(x_0)$:

$$\delta\tilde{\xi}(x_0) = i\omega\delta\tilde{n}\int \frac{dq}{(2\pi)}\frac{1}{-i\omega + D_1q^2 + \kappa_-F^{-1}(q)} \tag{61}$$

$$F(q) = 1+\frac{\kappa_+}{4\pi D_3}\log\left\{1+\frac{\Lambda^2}{-i\omega/D_3 + q^2}\right\}. \tag{62}$$

Compared to the pure 1D result, Eq (36), Eq (61) now contains a new term $\kappa_-F^{-1}(q)$ in the denominator; for $\kappa_- = 0$ this term vanishes and the result reverts to the 1D case as expected. $F(q)$ depends on the momentum q, whereas in the perfect mixing case, Eq (47), it was simply equal to 1. The integrand of Eq (61) is still divergent for $\omega = 0$ as $q \to 0$, but the integral nevertheless converges. Assuming a non-zero κ_-, the limit $\omega \to 0$ therefore exists, and we can integrate:

$$\delta\tilde{\xi}(x_0, \omega \to 0) = \frac{i\omega\delta\tilde{n}}{\pi\Lambda D_1}I(\alpha,\beta), \tag{63}$$

where

$$I(\alpha,\beta) = \int_0^\infty \frac{dt}{t^2 + \beta\left(1 + \alpha\log\left(1+t^{-2}\right)\right)^{-1}}, \tag{64}$$

$$\alpha = \frac{\kappa_+}{4\pi D_3}, \tag{65}$$

$$\beta = \frac{\kappa_-}{\Lambda^2 D_1} = \left(\frac{R}{\pi b}\right)^2. \tag{66}$$

The noise power spectrum for occupancy follows in close analogy to Eq (33):

$$S_n(\omega \to 0) = \frac{2\bar{n}(1-\bar{n})^2}{k_-} + \frac{2\bar{n}^2(1-\bar{n})^2}{\pi\Lambda D_1\bar{\xi}}I(\alpha,\beta). \tag{67}$$

The effective spectrum of local concentration fluctuations corresponding to the second term is (cf. Eqs (34, 38, 50)):

$$S_c(\omega \to 0) = \frac{2\bar{\xi}R}{\pi^2 D_1}I(\alpha,\beta). \tag{68}$$

The associated noise variance is:

$$\left(\frac{\sigma_\xi}{\bar{\xi}}\right)^2 = \frac{2I(\alpha,\beta)}{\pi^2 D_1\tau_{\text{int}}\bar{\xi}/R}. \tag{69}$$

With the above definition of α, the equilibrium between 1D and 3D concentrations turns out to be $\bar{c} = \bar{\xi}/4\pi D_3\alpha\tau_r$. The noise variance in concentration can be rewritten as:

$$\left(\frac{\sigma_c}{\bar{c}}\right)^2 = \frac{\beta I(\alpha,\beta)}{2\pi\alpha} \times \frac{1}{D_3\bar{c}\tau_{\text{int}}} \times \frac{1}{R} \tag{70}$$

The result looks similar to the pure 3D diffusion case in Eq (35): the noise due to concentration fluctuations has its length scale a (receptor size in pure 3D case) replaced with R (effective DNA cross-section in the combined 1D/3D case), and the noise contribution term gets multiplied by the prefactor that we examine next.

We expect the parameter $\alpha = \kappa_+/4\pi D_3$ to be close to 1 for diffusion limited approach to DNA. Imagine that the area that TF attempts to hit in order to bind non-specifically to the DNA is a cylindrical segment of DNA with radius R and length b. If we were treating the
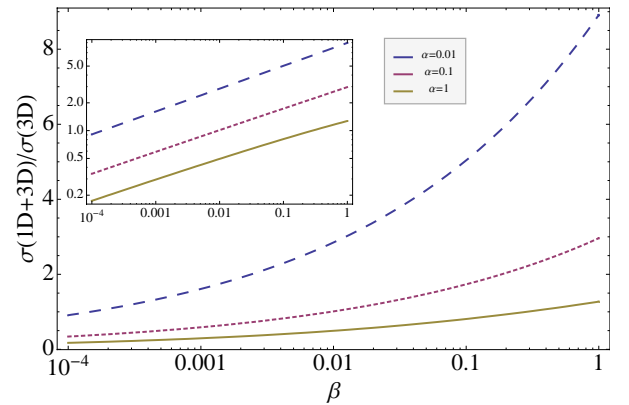

FIG. 2: The relative decrease in noise magnitude, compared to the pure 3D diffusion model, as a function of parameters α and β. Three values for α are shown, spanning two orders of magnitude ($\alpha = 1$ solid, $\alpha = 0.1$ dotted, $\alpha = 0.01$ dashed); β covers the relevant range if typical 1D diffusion length is as expected from search time optimality arguments (order hundred base pairs).

cylindrical DNA segment as a sphere of radius b, then Smoluchowski on-rate limit $\sim 4\pi D_3 b$ would apply [20]. In the first approximation, the on rate per unit length, κ_+, would then be $4\pi D_3$ and $\alpha = 1$. Exact derivations for Smoluchowski limit in simple geometries were reviewed in Ref [21]; for elongated objects (prolate ellipsoid with long semiaxis b and short semiaxis R), the diffusion limited rate is given by $4\pi D_3 b / \ln(2b/R)$, and therefore $\alpha \sim \ln^{-1}(2b/R)$. α is hence less then one, and has a weak logarithmic dependence on the ratio of the scale of the persistence length and the effective radius of the DNA.

The parameter β is approximately the square of the ratio between the cross-section of the 1D cylinder (the "target" that 3D diffusion has to hit) and the average "sliding length" b along the DNA. R must be of order of several nanometers; while b is, at DNA stacking length of $a = 0.3$ nm per base-pair and 100 bp average diffusion length [4, 14], around $b = 10 - 100$ nm. It is therefore not unreasonable to assume that the factor β could be as low as $\beta \sim 10^{-3} - 10^{-2}$. The corresponding decrease in noise relative to pure 3D diffusion is shown in Fig 2.

Note that as $\beta \to \infty$ one should recover the pure 3D diffusion result. Looking back at expression for $I(\alpha, \beta)$ in Eq (64), as β increases, the t^2 term in the denominator of the integrand is becoming irrelevant. If we neglect it completely, the integral is solvable analytically:

$$I(\alpha, \beta \to \infty) = \beta^{-1}\left(\frac{\Lambda_x}{\Lambda} + \pi\alpha\right), \qquad (71)$$

where $\Lambda_x = \frac{\pi}{a}$ is the cut-off at the binding site size along the DNA, and $\Lambda_x/\Lambda = R/a$. Inserting the integral in the large β limit into the noise result, Eq (70), we see that β cancels and we get:

$$\left(\frac{\sigma_c}{\bar{c}}\right)^2 = \frac{1}{2\pi}\left(\pi + \frac{R}{\alpha\, a}\right) \times \frac{1}{D_3 \bar{c} \tau_{\text{int}}} \times \frac{1}{R} \quad (72)$$

$$> \frac{1}{2\pi\alpha} \times \frac{1}{D_3 \bar{c} \tau_{\text{int}}} \times \frac{1}{a}, \qquad (73)$$

which is essentially the 3D result, Eq (35).

VI. DISCUSSION

While the role of 1D diffusion of transcription factors along the DNA has been recognized as a possible explanation for the observed fast DNA-TF association rates, the impact of this additional mode of TF translocation on the noise in binding site occupancy remained unexplored. The question is important for two reasons: first, the diffusive contribution to the noise in gene expression must fundamentally limit the precision of transcriptional regulation; and second, there is an appealing argument that 1D diffusion could drastically increase the target area on the DNA that TFs have to find and correspondingly lower the limiting diffusive noise. Here we show

that this intuitive argument is wrong—while there might be some reduction in the noise if the bulk diffusion is supplemented by 1D sliding, this reduction is not expected to be significant.

Much has been said about possible TF translocation strategies on the DNA such as three-dimensional volume exchange, local dissociation-reassociation reactions (hopping), sliding along the DNA and intersegmental transfers mediated by DNA-looping [3, 22], and about the ways in which these mechanisms influence the target search times [4, 23, 24, 25]. While our model does not examine all of the proposed mechanisms and makes approximations about the geometry of the TF and the DNA, we are ultimately only interested in what happens locally around the specific site as opposed to computing global properties such as target search time statistics. As a result, while rare but long range excursions might be important in search arguments, they should not significantly affect the noise as long as they are not the dominant form of all bulk transfers; if they are, one would have the case discussed in Section V.A of perfect mixing upon dissociation. Moreover, the conclusions presented here reflect the basic differences in the diffusive processes in one and three dimensions, in particular the high probability of returning to the origin in 1D diffusion, and are therefore not a consequence of the detailed assumptions about the TF-DNA interaction, as is clear from the "back of the envelope" arguments presented in the Section II.

The diffusive contribution to the total noise in gene expression has only recently been recognized as significant. Metzler discusses the concept of "interaction volume" around the regulatory site on the DNA and studies the probability that transcription factors will enter this volume in the case of the λ-phage infecting *Escherichia coli* bacterium, concluding that the spatial fluctuations can be important in genetic circuits [16]. Holcman et al study the master-diffusion equation in the context of signaling molecules binding to and unbinding from ion channels [17]. The analysis of Bicoid/Hunchback system in the development of the fruit fly *Drosophila melanogaster* shows that the measured signatures in the noise in Hunchback expression are consistent with diffusive fluctuations in Hunchback's regulator, Bicoid [18, 19]. Van Zon and coworkers [15] study the diffusional component of the noise in Green's function reaction dynamics (GFRD) stochastic simulations and conclude by noting that 1D sliding along the DNA could have important effects on the noise power spectra. All these results point to the basic physics of diffusion as setting a limit to precision of a fundamental biological process, transcriptional regulation. Our results indicate that this fundamental limit is not easily evaded by the (still largely unknown) complexities of protein motion along the DNA molecule.

[1] This limit is analogous to that identified long ago in the context of bacterial chemotaxis: HC Berg & EM Purcell, Physics of chemoreception, *Biophys J* **20**, 193–219 (1977). For a generalization of the Berg–Purcell argument, see Refs [11, 12]. For evidence that real transcriptional control systems approach this limit, see Refs [18, 19].

[2] AD Riggs, S Bourgeois & M Cohn, The *lac* repressor–operator interaction. III: Kinetic studies. *J Mol Biol* **53**, 401–417 (1970).

[3] OG Berg, RB Winter & PH von Hippel, Diffusion–driven mechanisms of protein translocation on nucleic acids: 1. Models and Theory. *Biochem* **20**, 6929–6948 (1981).

[4] M Slutsky & LA Mirny, Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys J* **87**, 4021–4035 (2004).

[5] DM Gowers, GG Wilson & SE Halford, Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Nat'l Acad Sci USA* **102**, 15883–15888 (2005).

[6] H Kabata, O Kurosawa, I Arai, M Washizu, SA Margarson, RE Glass & N Shimamoto, Visualization of single molecules of RNA polymerase sliding along DNA. *Science* **262**, 1561–1563 (1993).

[7] YM Wang, RH Austin & EC Cox, Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys Rev Lett* **97**, 048302 (2006).

[8] J Elf, GW Li & XS Xie, Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).

[9] Several colleagues have presented this intuitive argument as a response to the discussion in Ref [18], which claims that the observed precision of transcriptional regulation in the early fruit fly embryo is close the physical limit. We don't know of anyone who has put this argument into print, however. We take this opportunity to thank those who insisted on a more rigorous answer.

[10] EM Lifshitz & LP Pitaevskii, *Statistical Physics Part 2*. JB Sykes & MJ Kearsley, trans. (Pergamon Press, Oxford, 1980). See Section 89.

[11] W Bialek & S Setayeshgar, Physical limits to biochemical signaling. *Proc Nat'l Acad Sci USA* **102**, 10040–10045 (2005).

[12] W Bialek & S Setayeshgar, Cooperativity, sensitivity and noise in biochemical signaling. *arXiv.org:q–bio.MN/0601001* (2006).

[13] In case of transcriptional control, where the regulated protein is the "readout" of the transcription factor concentration, the integration time is usually the regulated protein lifetime (minutes or more); during this time the specific site occupancy will have equilibrated, implying that many translocation-related steps (binding and unbinding transitions to/from the specific site and the sliding state) occur during one integration time.

[14] SE Halford & JF Marko, How do site–specific DNA–binding proteins find their targets? *Nucl Acids Res* **32**, 3040–3052 (2004).

[15] JS van Zon, MJ Morelli, S Tanase-Nicola & PR ten Wolde, Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys J* **91**, 4350–4367 (2006).

[16] R Metzler, The future is noisy: the role of spatial fluctuations in genetic switching. *Phys Rev Lett* **87**, 068103-1 (2001).

[17] D Holcman & Z Schuss, Stochastic chemical reactions in micro-domains. *arxiv.org:math-ph/0412089v1* (2001).

[18] T Gregor, DW Tank, EF Wieschaus & W Bialek, Probing the limits to positional information. *Cell* **130**, 153–164 (2007).

[19] G Tkacik, T Gregor & W Bialek, The role of input noise in transcriptional regulation. *arXiv.org:q-bio/0701002* (2007).

[20] Diffusion-limited reaction rates generally take the form $k_+ = 4\pi DL$, where L is the effective target size that is primarily determined by the largest *linear* extent of the target. See Ref [21] for details.

[21] OG Berg & PH von Hippel, Diffusion-controlled macromolecular interactions. *Ann Rev Biophys Chem* **14**, 131–160 (1985).

[22] N Shimamoto, One-dimensional diffusion of proteins along DNA. *J Biol Chem* **274**, 15293–15296 (1999).

[23] IM Sokolov, R Metzler, K Pant & MC Williams, Target search of N sliding proteins on a DNA. *Biophys J* **89**, 895–902 (2005).

[24] M Coppey, O Bénichou, R Voituriez & M Moreau, Kinetics of target site localization of a protein on DNA: a stochastic approach. *Biophys J* **87**, 1640-1649 (2004).

[25] MA Lomholt, T Ambjörnsson & R Metzler, Optimal target search on a fast-folding polymer chain with volume exchange. *Phys Rev Lett* **95**, 260603 (2005).